DTU Library

Analysis of the results of an on-line wind power quantile forecasting system

Nielsen, H.Aa.; Yates, D.; Madsen, H.; Nielsen, T.S.; Badger, Jake; Giebel, Gregor; Landberg, Lars; Sattler, K.; Feddersen, H.

Publication date:
2006

Document Version
Publisher's PDF, also known as Version of record

PSO (FU 2101)
Ensemble-forecasts for wind power

Analysis of the Results of an On-line Wind Power Quantile Forecasting System

Henrik Aalborg Nielsen[1], Devon Yates[1],
Henrik Madsen[1], Torben Skov Nielsen[1],
Jake Badger[2], Gregor Giebel[2], Lars Landberg[2],
Kai Sattler[3], Henrik Feddersen[3]

[1] Technical University of Denmark, Informatics and Mathematical Modelling, DK-2800 Lyngby
[2] Risø National Laboratory, Wind Energy Department, DK-4000 Roskilde
[3] Danish Meteorological Institute, Research and Development, DK-2100 København Ø

10th October 2005

Contents

1 Summary 2

2 Introduction 2

3 Data 5

4 Reliability 8

5 Sharpness 12

6 Resolution 14

7 Spread / skill relationship 16

8 Conclusion and Discussion 19

9 Acknowledgements 20

References 20

A Plots related to reliability 22

B Sharpness and Resulution for a range of quantiles 28

1 Summary

Based on models and methods developed as part of the Danish PSO project *Ensemble-forecasts for wind power* (FU2101) a demo-application producing quantile forecasts of wind power has been developed. The application has been used by Elsam Kraft A/S and Energi E2 A/S, i.e. the two utilities participating in the project. The quality of the produced forecasts are addressed in this report.

Based on the results presented in the report it is shown that approximately reliable (i.e. probabilistic correct) quantile forecasts of the wind power production are generated by the application. However, this requires that the application is re-calibrated on a regular basis. Here a monthly re-calibration gave satisfactory results.

The results obtained using either ECMWF (European Centre for Medium-Range Weather Forecasts) or NCEP (National Centers for Environmental Prediction in the U.S.) ensemble forecasts of wind speed and direction are similar. However, for the setup used by Elsam Kraft A/S the reliability of the quantile forecasts when using NCEP ensembles is not quite satisfactory. Also, with respect to sharpness the two types of meteorological ensemble forecasts result in similar performance. However, for horizons between 36 and 48 hours the ECMWF ensemble seems to result in slightly sharper forecasts, i.e. more precise forecasts, than when using the NCEP ensembles. This behaviour is consistent for both setups.

Finally, it is shown that when considering the 50% quantile (median) forecast as a point forecast the magnitude of the error of this forecast is related to the difference between the 25% and the 75% quantile forecasts. This demonstrate the spread / skill relationship of the power quantile forecasting system. Furthermore, this relationship does not show any obvious dependence of the range of horizons considered. This further confirms that the uncertainty indicated by the quantile forecast system indeed reflects the true uncertainty of the point forecast.

2 Introduction

Based on models and methods developed as part of the Danish PSO project *Ensemble-forecasts for wind power* (FU2101) (Giebel et al., 2004) a demo-application producing quantile forecasts of wind power has been developed. The models and methods are described in (Nielsen et al., 2004) and (Nielsen et al., 2005). The application is hosted at IMM, via servers at DMI and Risø the application receives daily ensemble forecasts of 10m wind speed and direction from the European Centre for Medium-Range Weather Forecasts (ECMWF), where after the application produces quantile forecasts of the wind power production. A more detailed description of the ECMWF ensemble prediction system can be found in (Nielsen et al., 2004, 2005) and the references therein. The resulting quantile forecasts and wind power ensembles are available by the utilities via password protected homepages on the Internet. The setup for the two utilities participating in the project can briefly be described as:

- For Elsam Kraft A/S the total production in Jylland/Fynen, with the exception of Horns Rev, is forecasted on an hourly basis.
- For Energi E2 A/S the production at Nysted Offshore is forecasted at on a 15 minute basis.

The system whereby the ECMWF-ensembles are transformed into wind power ensembles and quantile forecasts are calibrated to actual power measurements. For each such calibration the latest six months are used. Figure 1 show an example of a quantile forecast.

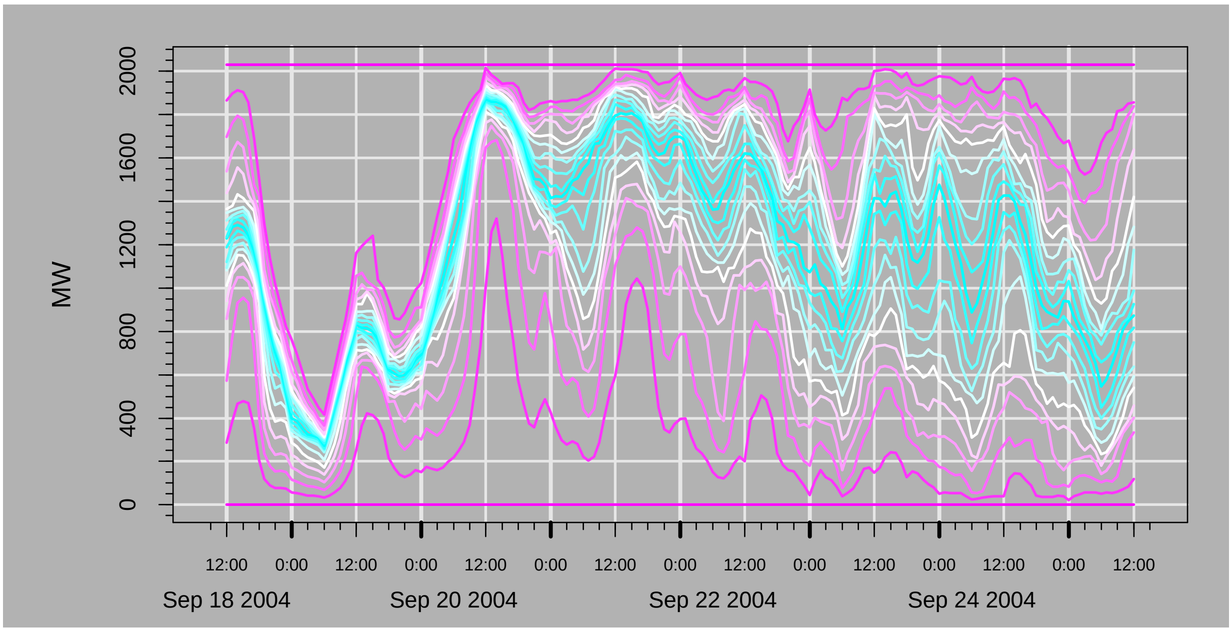

Figure 1: Example of a quantile forecast for the Elsam setup. The forecast is based on the ECMWF ensemble forecast for which the calculations were initiated at Sept. 18, 2004, 12:00 (UTC). The forecast were ready for the plan submitted to NordPool on Sept. 19, 2004. The 50% quantile is indicated by cyan and quantiles further away from the 50% quantile is color coded going over white to violet for the 0% and 100% quantiles.

It was planned that the calibration should have been performed every month. However, the calibration is not performed automatically by the system and it was not calibrated on a regular basis. This was due to some difficulties within the chain of activities involved in the supply of power data to the re-calibration of the model. For this reason this reports also consider the results of a so called 'rerun' in which the application were run in retrospect with monthly calibration using power observations and ensemble forecasts from the latest six months. In an actual implementation, an automated transfer of data would be set up, avoiding this kind of delay.

National Centers for Environmental Prediction (NCEP) in the U.S. offers an alternative ensemble forecasting system, see (Nielsen et al., 2005) and the references therein. A rerun of the application using the NCEP ensembles were also performed.

In summary this report considers the following:

Elsam: Original run for the Elsam Kraft A/S setup.

E2: Original run for the Energi E2 A/S setup.

Rerun Elsam: Rerun of the Elsam Kraft A/S with regular calibration.

Rerun E2: Rerun of the Energi E2 A/S with regular calibration.

NCEP Elsam: Rerun of the Elsam Kraft A/S with regular calibration and NCEP ensembles.

NCEP E2: Rerun of the Energi E2 A/S with regular calibration and NCEP ensembles.

The quality of the quantile forecasts are evaluated using four different criteria:

Reliability addresses whether quantile forecasts are indeed quantiles by comparing the nominal value of the quantile to the number of times the observation is actually below the quantile considered. Repeating this procedure for a range of quantiles allows a plot of the actual versus nominal values to be constructed. Ideally, the resulting line should be the line of identity. In order to perform this analysis the data must be grouped. In this report we group the data by ranges of horizons.

Sharpness measures the average uncertainty indicated by the quantile forecast system by addressing the difference between forecasted quantiles which are symmetric about the 50% quantile. The sharpness is calculated individually for each horizon and quantified both in terms of the mean and the median of the difference in quantiles.

Resolution measures the variation in uncertainty indicated by the quantile forecast system by addressing the same basic quantities as for sharpness, but by calculating measures of variation instead of e.g. the mean. The resolution is calculated individually for each horizon and quantified both in terms of the standard deviation (SD) and the median absolute deviation (MAD).

The spread / skill relationship address the relation between the magnitude of the error of a point forecast, i.e. a forecast consisting of a single value for each horizon / forecast time, and the uncertainty indicated by the quantile forecast system. For this analysis there should be a *tendency* of larger errors when the uncertainty indicated by the quantile forecast system is high. Note that it is the *spread* of the errors which should be reflected by the quantile forecast system.

More details about the data are listed in Section 3, which also describes the calibration dates for the actual use of the demo application. In Section 4 the reliability of the ensemble forecasts are considered. Note that in (Nielsen et al., 2004, 2005) QQ-plots are used to judge reliability. In this report we use reliability diagrams which simply displays the actual frequency against the nominal frequency, see also (Pinson et al., 2006). Sharpness, resolution and the spread / skill relation is considered in Sections 5, 6, and 7 respectively. Note that the relevant quantities is calculated in all cases even though these are only relevant for reliable quantile forecast systems. Finally, in Section 8 we conclude on the results.

3 Data

The measured power and ensemble forecasts used in the investigation are characterized by:

Measured Power: Real power measurements are taken at two different wind sites, referred to as E2 and Elsam. The specific measurement sites are:

- *E2* 15 minute mean power from E2's Nysted Offshore wind farm. Installed capacity: 165.6 MW[1].
- *Elsam* 1 hour mean power measurements from the combined production of all wind parks in Jylland and Fynen (Eltra area), excluding Horns Rev. Installed capacity: 2216 MW[2]

ECMWF Ensemble Forecasts: Wind speed and direction 10 meters above ground level from the ECMWF ensemble forecasting system are extracted from the archive at Risø for the 51 members and applied to the demonstration locations for E2 and Elsam. The temporal resolution of the forecast is 6 hours, and the horizon is 7 days. The forecast cycles begin at 12Z each day. The spatial resolution is approximately 75 km. The wind data is applied to the site in question in the following manner,

- *E2* The nearest 4 grid points (11.25E 54.00N, 12.00E 54.00N, 11.25E 54.75N, 12.00E 54.75N) from the forecast are used for bilinear interpolation to the wind farm location (11.7597E 54.54075N) of u and v (westerly and southerly component of the wind) and from these absolute wind speed and direction is calculated.

[1] Actually depending on the configuration of the wind farm, but here we use the number of turbines times the capacity of one turbine (2.3 MW).

[2] Actually ranging from 2214 to 2218 MW for the period considered.

Elsam The mean of u and v (westerly and southerly component of the wind) is calculated using the 12 grid points covering Jylland and Fynen with coordinates (with weightings in brackets):

8.25E 57.00N (1/24)	9.00E 57.00N (1/12)	9.75E 57.00N (1/12)	10.5E 57.00N (1/24)
8.25E 56.25N (1/12)	9.00E 56.25N (1/6)	9.75E 56.25N (1/6)	10.5E 56.25N (1/12)
8.25E 55.50N (1/24)	9.00E 55.50N (1/12)	9.75E 55.50N (1/12)	10.5E 55.50N (1/24)

The mean u and v is used to calculate the mean absolute wind speed and direction for the region.

NCEP Ensemble Forecasts: Wind speed and direction 10 meters above ground level from the NCEP ensemble forecasting system are extracted from archive at Risø for the 11 members and applied to the demonstraton locations for E2 and Elsam. The temporal resolution of the forecast is 6 hours, and the horizon is 7.5 days (3.5 days before 01/04/2004). Only the 00Z forecast cycle is used, see (Nielsen et al., 2005). The spatial resolution is approximately 100 km. The wind data is applied to the site in question in the following manner,

E2 The nearest 4 grid points (11.0E 54.00N, 12.00E 54.00N, 11.0E 55.0N, 12.00E 55.0N) from the forecast are used for bilinear interpolation to the wind farm location (11.7597E 54.54075N) of u and v (westerly and southerly component of the wind) and from these absolute wind speed and direction is calculated.

Elsam The mean of u and v (westerly and southerly component of the wind) is calculated using the 12 grid points covering Jylland and Fynen with coordinates (with weightings in brackets):

8.0E 57.0N (1/24)	9.0E 57.0N (1/12)	10.0E 57.0N (1/12)	11.0E 57.00N (1/24)
8.0E 56.0N (1/12)	9.0E 56.0N (1/6)	10.0E 56.0N (1/6)	11.0E 56.0N (1/12)
8.0E 55.0N (1/24)	9.0E 55.0N (1/12)	10.0E 55.0N (1/12)	11.0E 55.0N (1/24)

The mean u and v is used to calculate the mean absolute wind speed and direction for the region.

Temporal Interpolation of Ensemble Forecasts: For both types of ensemble forecasts linear interpolation is used in order to assign forecasts to the time points of all measured power values.

Calibration of actual application: The quantile prediction for the E2 site was begun on June 28, 2004, and for the Elsam data on June 12, 2004. Each site system was retrained periodically, as shown in Table 1. A graphical representation of the training schedule is shown in Figure 2 and 3, where the y-axis is the number of times that the system has been trained. It can be seen that the training is done out of sample. It is also noted that the prediction periods have some overlap because the horizon of the predictions extends quite far ahead. Additionally, because the retraining was done manually, there was some lag between end of the training, and the initiation of the predictions.

Period: The comparisons between actual runs and reruns are performed based on the same time-period in all cases. Considering the time points at which the model calculations were initiated at the meteorological centres the resulting period start at July 28, 2004 and end at March 31, 2005, i.e. a total of 8 months.

Site	Prediction Range	Train Start	Train End
Elsam	2004/07/12 12:00 - 2005/02/03 12:00	2004/01/01 12:00	2004/05/31 12:00
Elsam	2005/02/03 12:00 - 2005/03/31 12:00	2004/07/01 12:00	2004/12/31 12:00
E2	2004/07/28 12:00 - 2004/10/31 12:00	2004/03/27 23:00	2004/06/30 12:00
E2	2004/10/25 12:00 - 2005/02/04 12:00	2004/04/01 12:00	2004/09/30 12:00
E2	2005/02/04 12:00 - 2005/02/03 12:00	2004/07/01 00:00	2004/12/31 23:45
E2	2005/02/04 12:00 - 2005/03/31 12:00	2004/08/01 00:00	2005/01/31 23:00

Table 1: Prediction and Training Periods of the actual demo application.

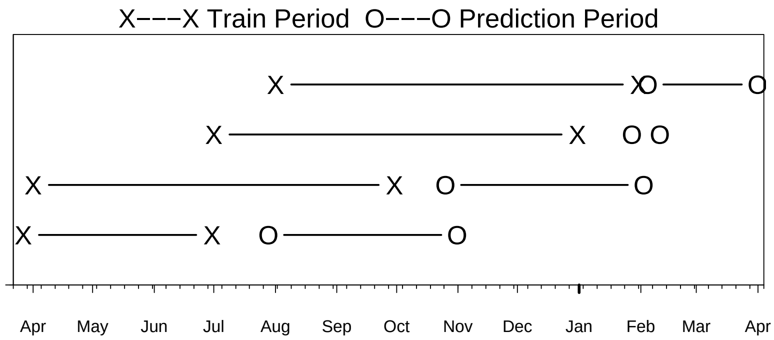

Figure 2: Sequence of Training and Prediction Periods for E2

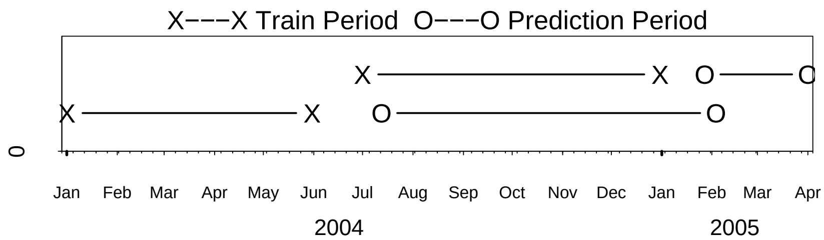

Figure 3: Sequence of Training and Prediction Periods for Elsam

4 Reliability

The ECMWF ensemble system is initialized daily at 12:00 (UTC). Taking into account the calculation time, but disregarding daylights savings and the one hour difference between UTC and the time zone in the NordPool area, the relevant horizons for bidding at the NordPool market is 36 to 60 hours. Figure 4 shows reliability diagrams for these horizons and Tables 2 and 3 show the corresponding numbers. These diagrams and tables are constructed by counting actual (relative) number of times the observation is below a quantile forecast with a given nominal value. Ideally, when the period under consideration is very long, there should be a complete agreement between the actual and nominal values. Unfortunately, for finite sized periods, it is not simple to quantify the deviation which can occur by sheer chance (Pinson et al., 2006). As seen from Figures 2 and 3 the re-calibration or training has been performed on a much more regular basis for E2 than for Elsam. This fact seem to be reflected in the reliability diagrams of the original runs; for E2 the line is much closer to the line of identity than for Elsam. For the rerun, i.e. the regular calibration, nearly the same curve is obtained for E2 and for Elsam the reliability is markedly improved when calibrated at regular basis. This highlights the importance of adaptive methods or at least regular re-calibration[3]. The rerun using the NCEP ensembles show a somewhat similar performance, but especially for Elsam the deviation from the line of identity is not quite satisfactory. In Appendix A reliability diagrams for horizons ranging from 18 to 168 hours in steps of six hours are displayed. From these plots similar conclusions are reached.

With the purpose of summarizing the reliability diagrams for the different horizons in a single number the Mean Squared Error for each diagram is calculated, i.e. the mean of the squared difference between the actual and nominal values is calculated for different horizons. Here we consider horizons up to 168 hours and perform the calculations for each full hour. The result is displayed in Figure 5. This confirms that for the Elsam setup there seems to be a benefit in using the ECMWF ensembles rather than the NCEP ensembles. For the E2 setup there seem to be a very small benefit of NCEP over ECMWF. Note also the difference in initialization time for NCEP and ECMWF which is seen for the Elsam setup, especially for the shorter horizons.

[3]Given daily updates of the power data this can easily be done since the re-calibration of the statistical models takes less than 10 minutes, when using six months of data.

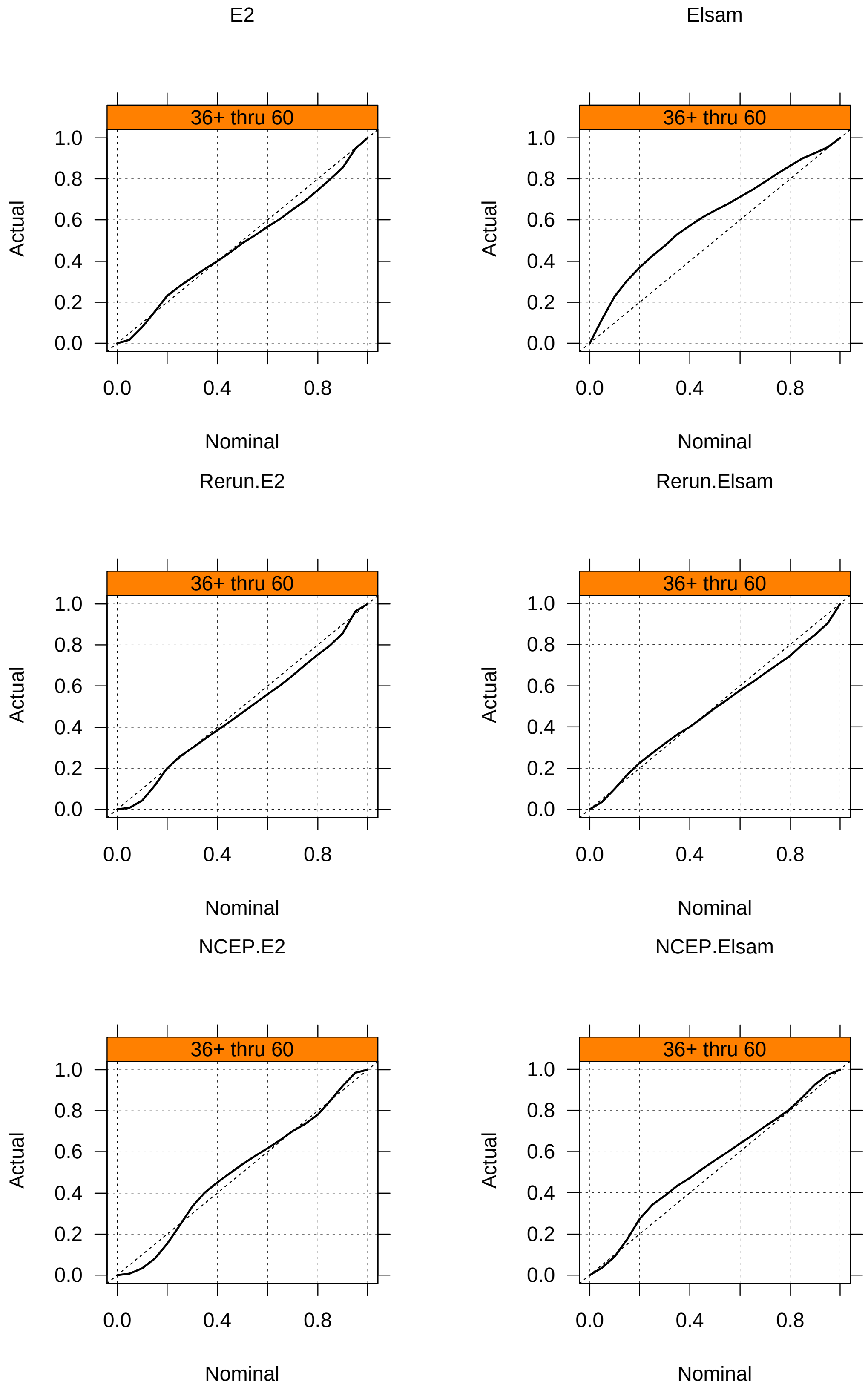

Figure 4: Reliability for horizon most relevant for trading on NordPool.

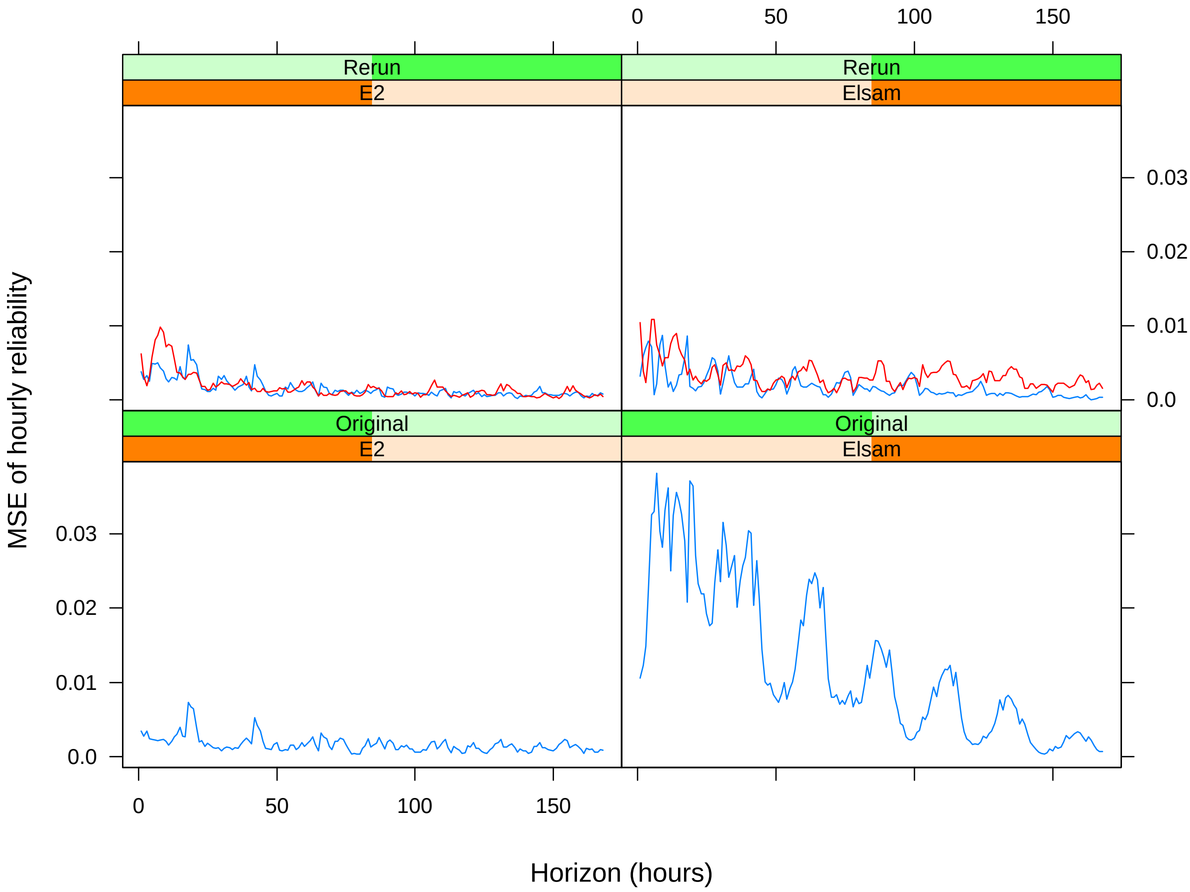

Figure 5: Mean squared error of the deviation of the reliability curve from the line of identity for ECMWF (blue) and NCEP (red).

Nominal	E2	Elsam	Rerun.E2	Rerun.Elsam	NCEP.E2	NCEP.Elsam
5	1.7	11.6	0.8	3.6	0.7	3.5
10	7.9	22.9	4.2	10.0	3.4	9.0
15	15.6	30.7	11.8	16.8	8.2	17.5
20	23.0	36.8	19.9	22.4	15.3	27.2
25	27.8	42.5	25.8	27.3	24.3	34.1
30	32.2	47.6	30.0	32.1	33.5	38.7
35	36.2	53.0	34.3	36.3	40.2	43.3
40	40.0	57.4	38.4	40.1	45.2	47.1
45	44.3	61.4	42.7	44.5	49.5	51.5
50	48.6	64.7	47.2	48.9	53.9	55.7
55	52.6	67.6	51.6	53.2	58.0	59.7
60	56.9	71.2	56.1	57.8	61.9	63.9
65	60.6	74.9	60.3	61.9	65.9	68.1
70	65.0	78.6	65.2	66.0	70.1	72.2
75	69.4	82.8	70.4	70.3	73.7	76.1
80	74.6	86.5	75.2	74.6	78.2	80.9
85	79.9	90.0	79.8	80.0	84.7	86.4
90	85.5	92.6	85.7	84.8	92.3	92.5
95	94.9	95.6	96.4	90.4	98.6	97.4

Table 2: Norminal and actual probabilities in percent for horizon most relevant for trading on NordPool.

Nominal	E2	Elsam	Rerun.E2	Rerun.Elsam	NCEP.E2	NCEP.Elsam
5	-3.3	6.6	-4.2	-1.4	-4.3	-1.5
10	-2.1	12.9	-5.8	0.0	-6.6	-1.0
15	0.6	15.7	-3.2	1.8	-6.8	2.5
20	3.0	16.8	-0.1	2.4	-4.7	7.2
25	2.8	17.5	0.8	2.3	-0.7	9.1
30	2.2	17.6	0.0	2.1	3.5	8.7
35	1.2	18.0	-0.7	1.3	5.2	8.3
40	0.0	17.4	-1.6	0.1	5.2	7.1
45	-0.7	16.4	-2.3	-0.5	4.5	6.5
50	-1.4	14.7	-2.8	-1.1	3.9	5.7
55	-2.4	12.6	-3.4	-1.8	3.0	4.7
60	-3.1	11.2	-3.9	-2.2	1.9	3.9
65	-4.4	9.9	-4.7	-3.1	0.9	3.1
70	-5.0	8.6	-4.8	-4.0	0.1	2.2
75	-5.6	7.8	-4.6	-4.7	-1.3	1.1
80	-5.4	6.5	-4.8	-5.4	-1.8	0.9
85	-5.1	5.0	-5.2	-5.0	-0.3	1.4
90	-4.5	2.6	-4.3	-5.2	2.3	2.5
95	-0.1	0.6	1.4	-4.6	3.6	2.4

Table 3: Norminal probabilities and the deviation of the actual probabilities from the norminal for horizon most relevant for trading on NordPool.

5 Sharpness

Here sharpness of a quantile forecast is defined as the average or median size of the Inter Quartile Range (IQR), i.e. the difference between the 1st and the 3rd quartile (the 25% and 75% quantiles). Figure 6 shows these numbers, based on values of IQR normalized using the installed capacity, for each horizon. Due to the issues regarding reliability the original runs will not be commented further here.

Comparing the E2 and Elsam setup it is seen that the Elsam setup results in sharper forecasts than the E2 setup. This is expectable since the Elsam setup covers a large geographic region where errors tend to cancel each other. Also, for the larger region, the typical power output changes much more smoothly than for the single farm at Nysted, and the extremes (full power or zero power) are reached much less frequently. Comparing the NCEP and the ECMWF ensembles these seem to have approximately the same sharpness for the Elsam setup. However for the horizons mainly relevant for NordPool the ECMWF ensembles seems to result in somewhat sharper forecasts, especially for the horizons 36 to 48 hours. For the E2 setup this effect is even more pronounced.

Figure 15 on page 28 show similar plots for the difference of other quantiles symmetric about the median. It is notable that, considering the horizons relevant for NordPool, the difference between the 90% and 10% quantile forecast is on average not more than 30% of the installed capacity when considering the Elsam setup. For the E2 setup the corresponding value is approximately 70%. A possible explanation is the difference in geographical regions considered in the two setups. Note also that even for the longest horizons the measures sharpness does not exceed 60% of the installed capacity for the Elsam setup. For the E2 setup the corresponding number is 90%.

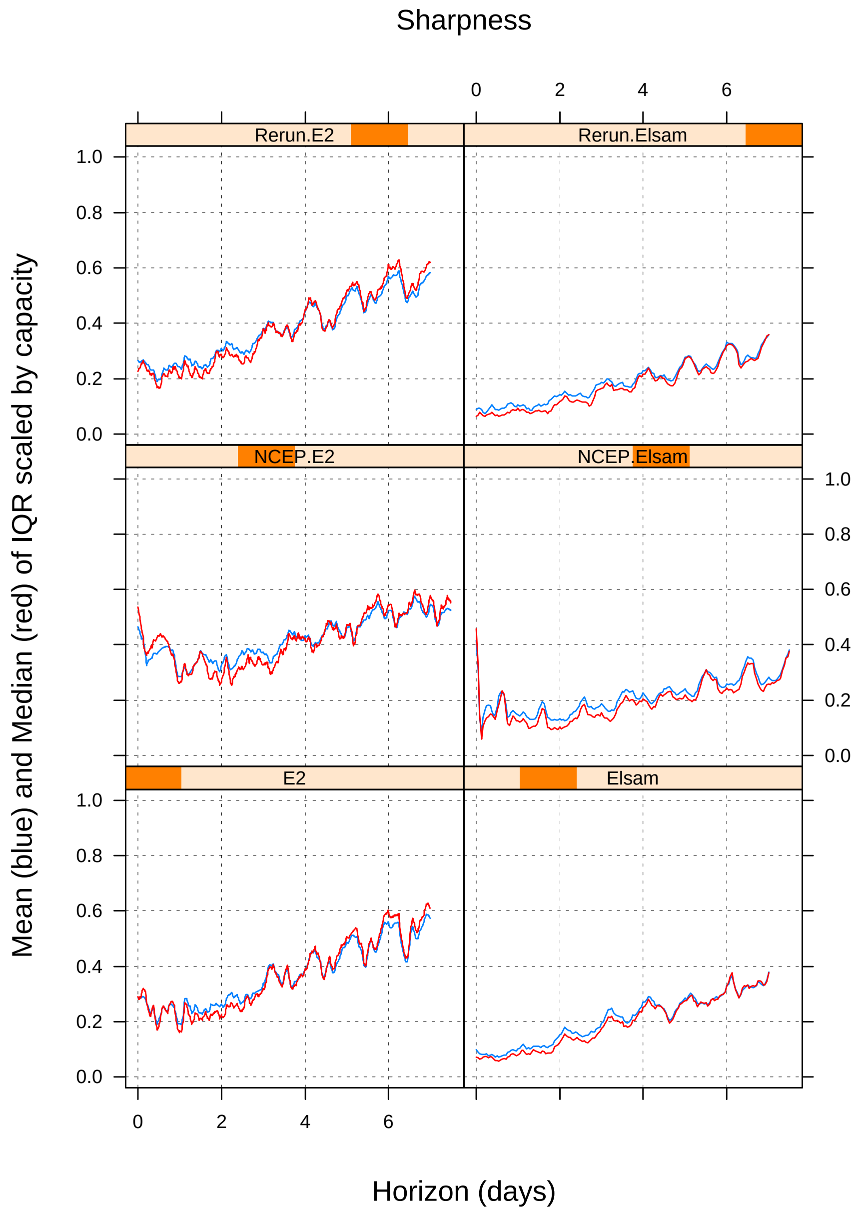

Figure 6: Sharpness compared to installed capacity.

6 Resolution

Here resolution of a quantile forecast is defined as the SD (Standard Deviation) and MAD (Median Absolute Deviation)[4] of the IQR. High values of SD and MAD indicates that the forecasting system is able to separate situations with high and low uncertainty. Given reliable systems with similar sharpness, the system with high resolution (high SD and MAD) may be preferable (assuming the resolution to be a real phenomenon and not originating from random noise on the estimates).

Figure 7 shows the values of SD and MAD, based on values of IQR normalized using the installed capacity, for each horizon. Due to the issues regarding reliability the original runs will not be commented further here.

First the Elsam setup is considered. Disregarding approximately the first 36 hours, where especially the NCEP-based quantile forecasts are not very reliable, it is seen that the systems have approximately the same resolution. The NCEP-based quantile forecasts seem to have slightly higher resolution than the ECMWF-based forecasts.

Considering the E2 setup the NCEP-based quantile forecasts seems to have slightly higher resolution than the ECMWF-based setup. Note that for horizons longer than four days the resolution for the E2-setup drops as the horizon increase. Comparing with the sharpness in Figure 6 it is seen that this happens when the average IQR higher than approximately half of the installed capacity. It is natural that in this case the variation in IQR can not continue to grow and it means that the uncertainty is often high for these horizons.

Comparing the Elsam and E2 setups w.r.t. resolution it is seen that the E2 setup has higher resolution than the Elsam setup. However, this is expectable since the Elsam setup produce markedly sharper forecasts than the E2 setup, cf. Figure 6, and in a strict sense the comparison is not appropriate.

Figure 16 on page 29 show similar plots for the difference of other quantiles symmetric about the median. Comparing with Figure 15 it is seen that the resolution decrease when the average distance (sharpness) approach the installed capacity.

[4]The median absolute deviation is multiplied by 1.4826, whereby it is approximately equal to the standard deviation for large Gaussian samples.

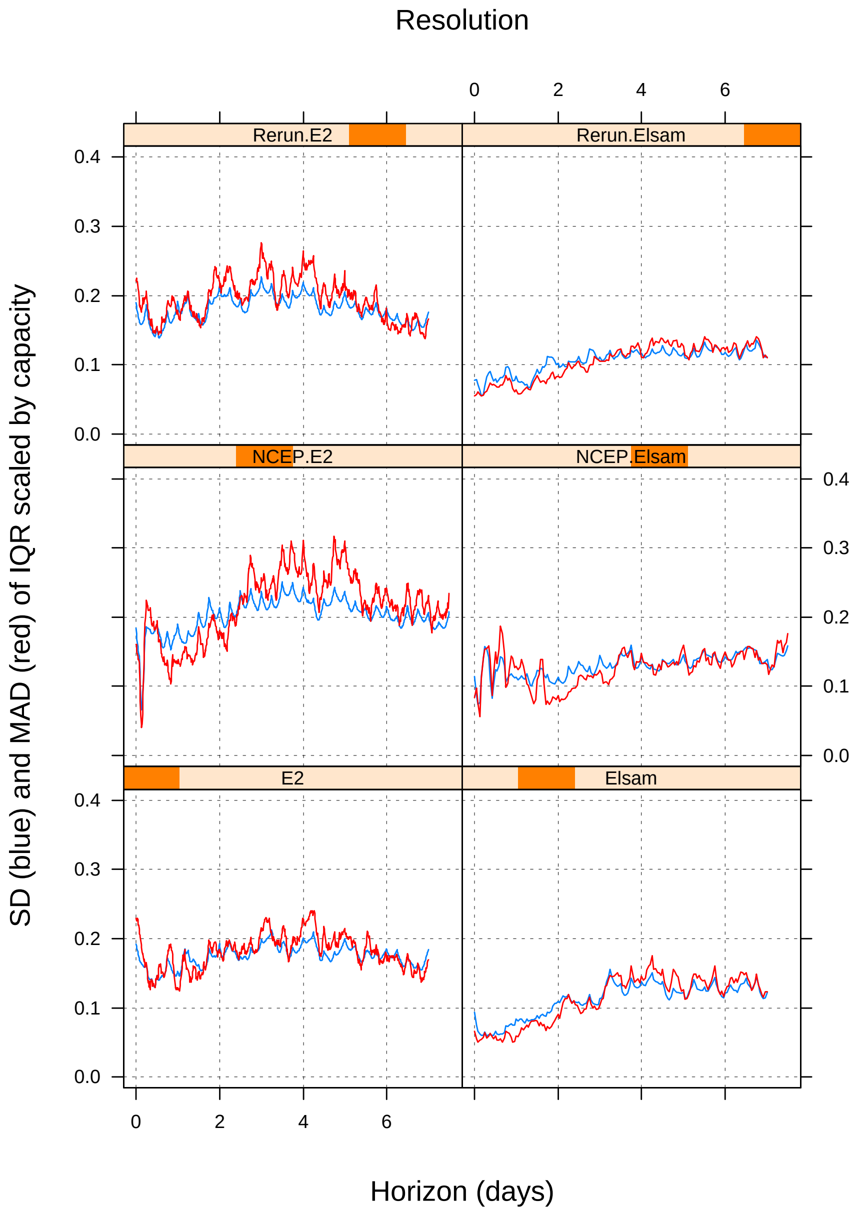

Figure 7: Resolution compared to installed capacity.

7 Spread / skill relationship

The spread / skill relationship refers to the relation between a point forecast (i.e. a single value) and the uncertainty indicated by the ensemble forecasting system. Here we will let the IQR mentioned in the previous sections represent the uncertainty indicated by the forecasting system. Several possibilities exists w.r.t. the point forecast:

- Use the control forecast from the ensemble system, i.e. the forecast without any perturbation of the initial values.
- Use the ensemble mean, i.e. the average of the power ensembles. This is similar to what is often done in meteorology, but the approach does not take into account that the power ensemble may not be probabilistic correct.
- Use the mean based on the corrected quantiles, i.e. construct the mean from the frequency distribution given by the quantiles. Given probabilistic correct quantiles this should give a more correct estimate of the mean than the ensemble mean. However, the extreme quantiles are not well defined from data and these may receive too much weight using this approach.
- Base the mean on the central, say 50%, part of the frequency distribution. This is similar to the above, but is not sensitive to errors in the extreme quantiles.
- Use the median forecast from the corrected quantiles. In terms of the Root Mean Square of the forecast errors (Madsen et al., 2004) this point forecast will probably not perform well compared to the mean. However, optimal bids on the spot market (NordPool) should often be a quantile close to the median (Bremnes, 2004). The cost criterion considered in (Nielsen and Ravn, 2003) also results in a bid on the spot market which is a quantile close to the median (the deviation from the median depends on the ratio of up- and down-regulation costs).

In this case it turns out that the mean based on the central part of the frequency distribution and the median results in similar values, especially for the Elsam setup. Since also the median seems to be more relevant from a market point of view the median forecast is selected as the reference.

Figure 8 shows the resulting spread / skill plots grouped by horizon. A clear relation between IQR and the magnitude of the absolute forecast error is evident from the plots. This clearly confirms that the quantile forecast systems indeed contains relevant information regarding the uncertainty. Note also that when disregarding the original runs and the horizons below 24 hours the plots are similar for a board range of horizons. This further confirms that the information regarding the uncertainty is contained in the values of IQR.

Note also that the purpose of the analysis of the spread / skill relationship is to confirm that the uncertainty forecasted by the quantile forecasts are indeed reflected in the uncertainty observed when using a point forecast. It seems logical to check this using the best point forecast available (given some criteria), disregarding if the point forecast is based on the quantile forecasting system. The best point forecast may indeed be a forecast based on a limited area model nested within the global model. Therefore it would make sense to repeat the spread / skill analysis using forecasts from e.g. WPPT (Madsen et al., 2005) as the point forecast.

Spread / Skill relationship

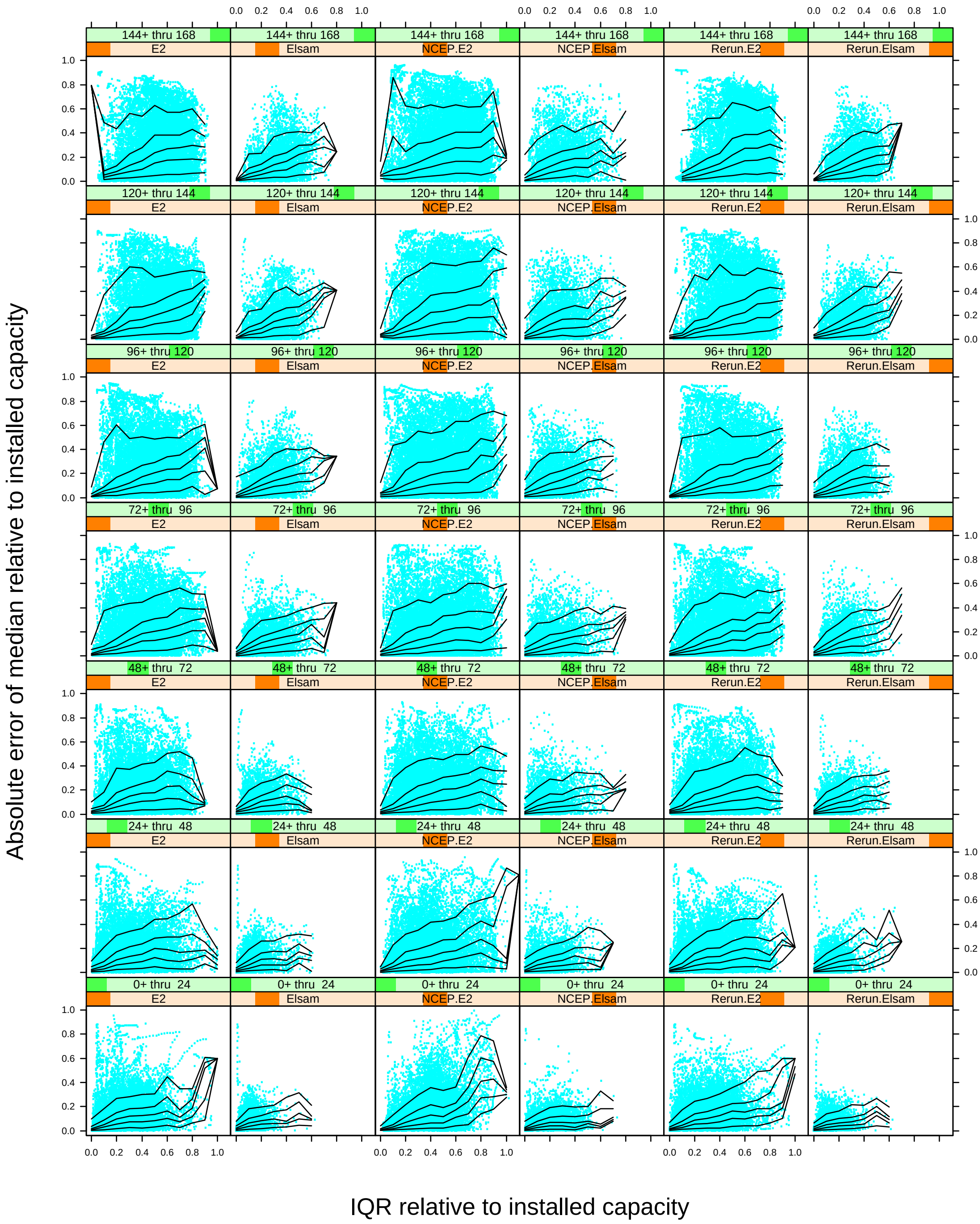

Figure 8: Spread / skill relationship when the 50% quantile forecast is used as the point forecast. The black lines indicate the lines for which 10%, 30%, 50%, 70%, and 90% of the errors are below, when grouping by the value on the 1st axis rounded to one decimal point.

8 Conclusion and Discussion

As described in the introduction a demo application of a quantile forecast system has been developed. This application has been used by the two utilities Elsam Kraft A/S and Energi E2 A/S participating in the project. In this report the results obtained during a period from summer 2004 until spring 2005 are analyzed. Furthermore, the results are compared with the results obtained had the system been re-calibrated on a more regular basis and had the ECMWF (European Centre for Medium-Range Weather Forecasts) ensemble forecasts been replaced with NCEP (National Centers for Environmental Prediction in the U.S.) ensemble forecasts.

From the analysis it is clear that regular re-calibration is extremely important. This also confirms that the use of adaptive methods are important for wind power forecasting in general, see e.g. the analysis regarding the optimal forgetting factor in (Nielsen, 1999). For a final software implementation the re-calibration should be performed automatically.

Given regular re-calibration it is shown that approximately reliable quantile forecasts can be produced. For the E2 setup (forecasts for Nysted Offshore) there is no large difference between the results when using the NCEP or the ECMWF ensemble forecasts. For the Elsam setup (forecast for all wind power in Jylland / Fynen except Horns Rev Offshore) the ECMWF ensemble forecasts seems to result in more reliable forecasts than when the NCEP-forecasts are used. In the probability model described in (Nielsen et al., 2004, 2005) a spline basis with only two internal knots is recommended. The same model and basis has been used for the demo application. Before an actual implementation it should however be investigated if the reliability of the quantile forecasts can be enhanced, without introducing excess variability, by increasing the number of knots. Furthermore, the bandwidth used in the probability model should be tailored to the specific setup.

The sharpness and resolution of the forecasts are analyzed. Comparing the results obtained using the ECMWF and the NCEP ensemble forecasts it is seen that approximately the same sharpness is obtained. However, especially for horizons between 36 and 48 hour the quantile forecasts based on the ECMWF ensembles is more sharp than the quantile forecasts based on the NCEP ensembles. Considering the calculation time of the models these horizons are very relevant from a market point of view. Note, however that the initialization times of the ECMWF and the NCEP models differ (12:00 and 00:00 (UTC), respectively) and therefore the comparisons outlined are mostly relevant from a scientific point of view. Also, w.r.t. resolution the difference in results when using the ECMWF or the NCEP ensembles is small. There is however a tendency for the quantile forecasts based on the NCEP ensembles to have a higher resolution than those based on the ECMWF ensembles.

When considering the difference between the 10% and 90% quantile forecasts it is notable that for the Elsam setup, which covers a large geographical area, the average difference between the quantiles is not more than 60% of the installed capacity for any horizon. For the horizons most relevant for bidding on NordPool the corresponding number is 30%. This confirms that on the large scale it is possible to forecast the wind power production to a large extend.

The spread / skill relation is also investigated. It is shown that there is a good relation between the error of the 50% quantile (median) forecast taken as a point forecast and the Inter Quartile Range (difference between the 25% and 75% quantile forecasts). Furthermore, this relation do not seem to be affected by the horizon considered. This indeed confirms that the actual uncertainty is reflected in the quantile forecasts.

As noted in (Nielsen et al., 2005) the estimation method used for the power curve model will not result in a central estimate of the power curve. The consequence will be that even if the ensembles were reflecting the true conditional probability distribution the bias of the estimated power curve would result in power ensembles which were not probabilistic correct. In turn this could result in large corrections performed by the probability model (Nielsen et al., 2005) and possibly a smaller sharpness than possible providing that unbiased estimates of the power curve could be found.

9 Acknowledgements

The project under which the work described here is carried out is sponsored by the Danish utilities PSO fund (ORDRE-101295 / FU 2101) which is hereby greatly acknowledged. Furthermore, the authors wish to thank the European Centre for Medium-Range Weather Forecasts, and the National Centers for Environmental Prediction in the U.S. for supplying the data used in this study.

References

Bremnes J.B. Probabilistic wind power forecasts using local quantile regression. *Wind Energy*, **7**(1):47–54, 2004.

Giebel G., Badger J., Landberg L., Nielsen H., Madsen H., Sattler K., and Feddersen H. Wind power forecasting using ensembles. In *Proceedings of the 2004 Global Windpower Conference and Exhibition*. Chicago, Illinois, USA, 2004. `http://www.awea.org`.

Madsen H., Kariniotakis G., Nielsen H., Nielsen T., and Pinson P. A protocol for standardizing the performance evaluation of short-term wind power prediction models. In *Proceedings of the 2004 Global Windpower Conference and Exhibition*. Chicago, Illinois, USA, 2004. `http://www.awea.org`.

Madsen H., Nielsen H.A., and Nielsen T.S. A tool for predicting the wind power production of off-shore wind plants. In *Proceedings of the Copenhagen Offshore*

Wind Conference & Exhibition. Danish Wind Industry Association, Copenhagen, 2005. http://www.windpower.org/en/core.htm.

Nielsen C.S. and Ravn H.F. Criteria in short-term wind power prognosis. In *Proceedings of the European Wind Energy Conference & Exhibition*. Madrid, Spain, 2003. http://www.ewea.org.

Nielsen H., Madsen H., Nielsen T., Badger J., Giebel G., Landberg L., Sattler K., and Feddersen H. Wind power ensemble forecasting. In *Proceedings of the 2004 Global Windpower Conference and Exhibition*. Chicago, Illinois, USA, 2004. http://www.awea.org.

Nielsen H.A., Madsen H., Nielsen T.S., Badger J., Giebel G., Landberg L., Sattler K., and Feddersen H. Wind power ensemble forecasting using wind speed and direction ensembles from ecmwf or ncep. Technical report, Informatics and Mathematical Modelling, 2005. Prepared under PSO-project FU2101.

Nielsen T. (editor). *Using Meteorological Forecasts in On-line Predictions of Wind Power*. Eltra, DK-7000 Fredericia, Denmark, 1999. ISBN: 87-90707-18-4.

Pinson P., Kariniotakis G., Nielsen H.A., Nielsen T.S., and Madsen H. Properties of quantile and interval forecasts of wind generation and their evaluation. In *Proceedings of the European Wind Energy Conference & Exhibition*. EWEA, Athens, 2006. http://www.ewea.org, to appear.

A Plots related to reliability

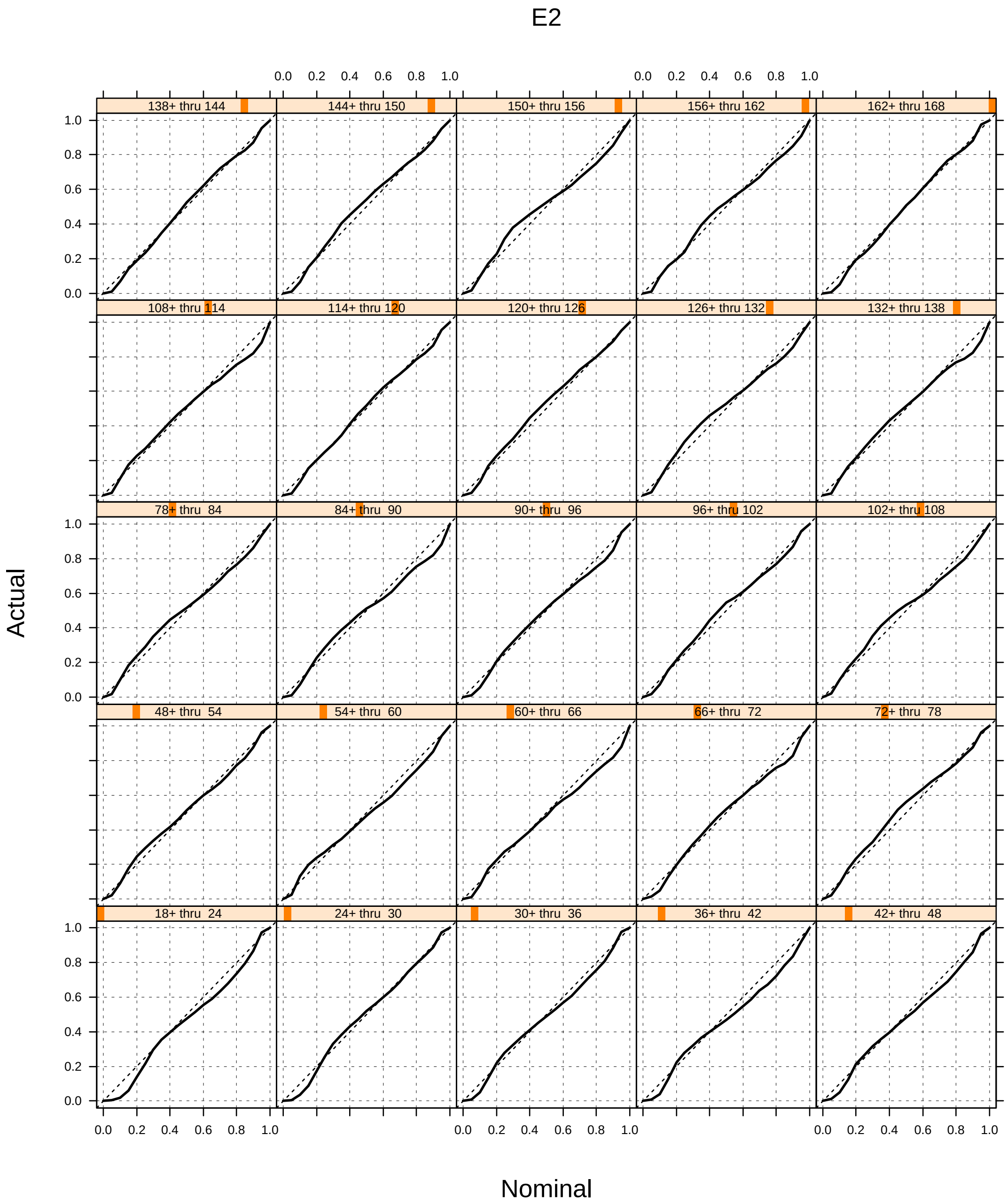

Figure 9: Reliability plots for E2.

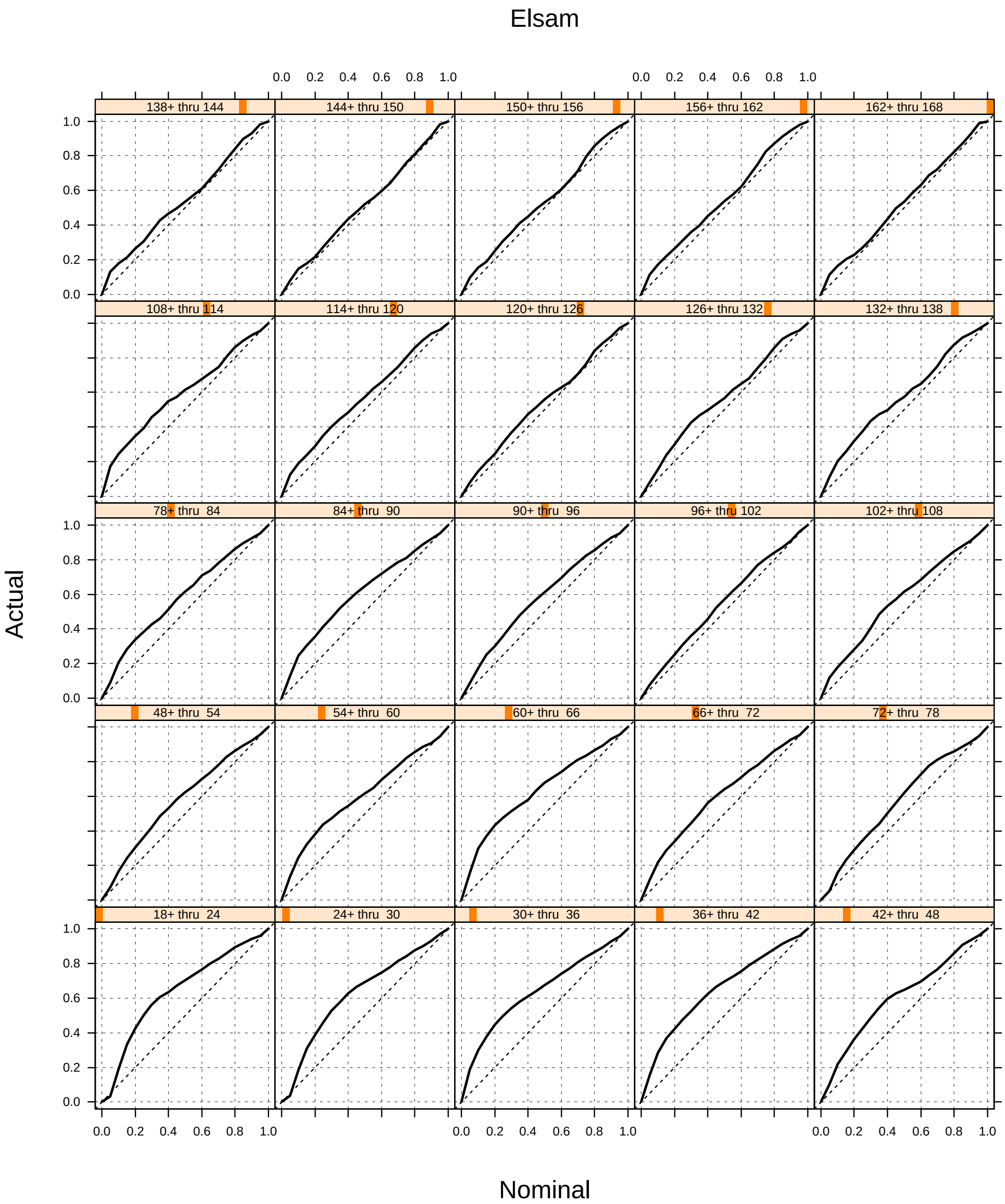

Figure 10: Reliability plots for Elsam.

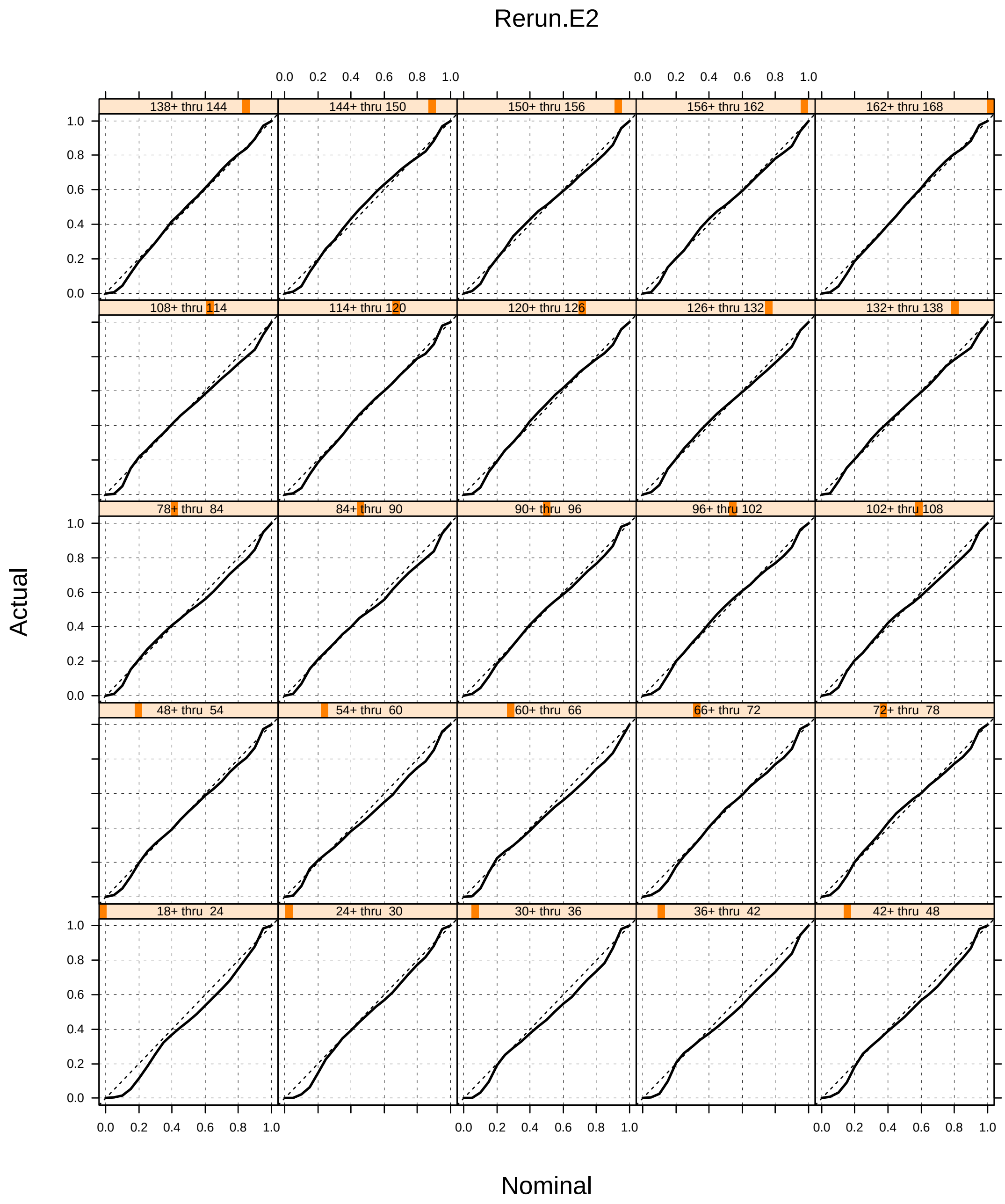

Figure 11: Reliability plots for rerun of E2 using ECMWF ensembles.

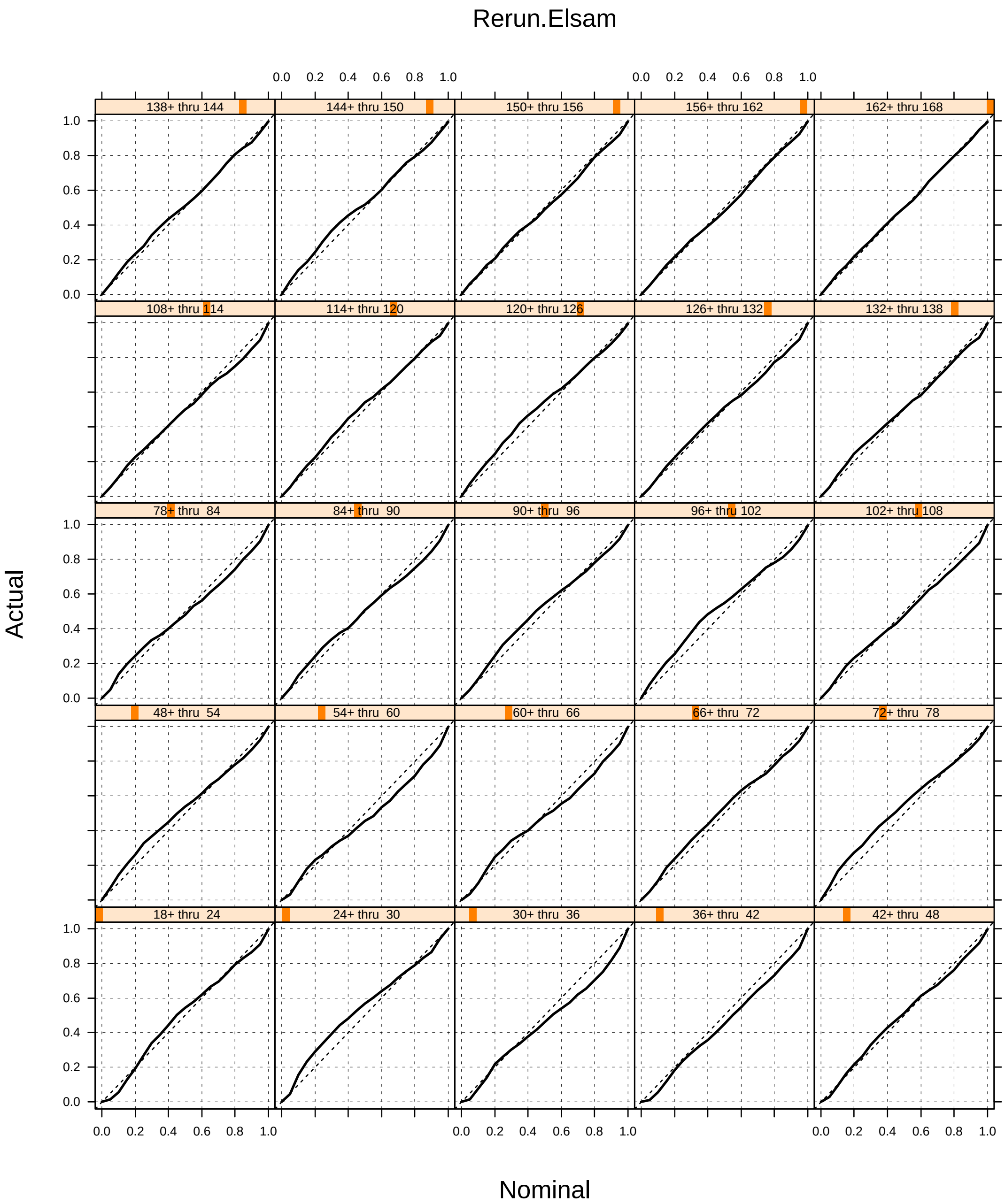

Figure 12: Reliability plots for rerun of Elsam using ECMWF ensembles.

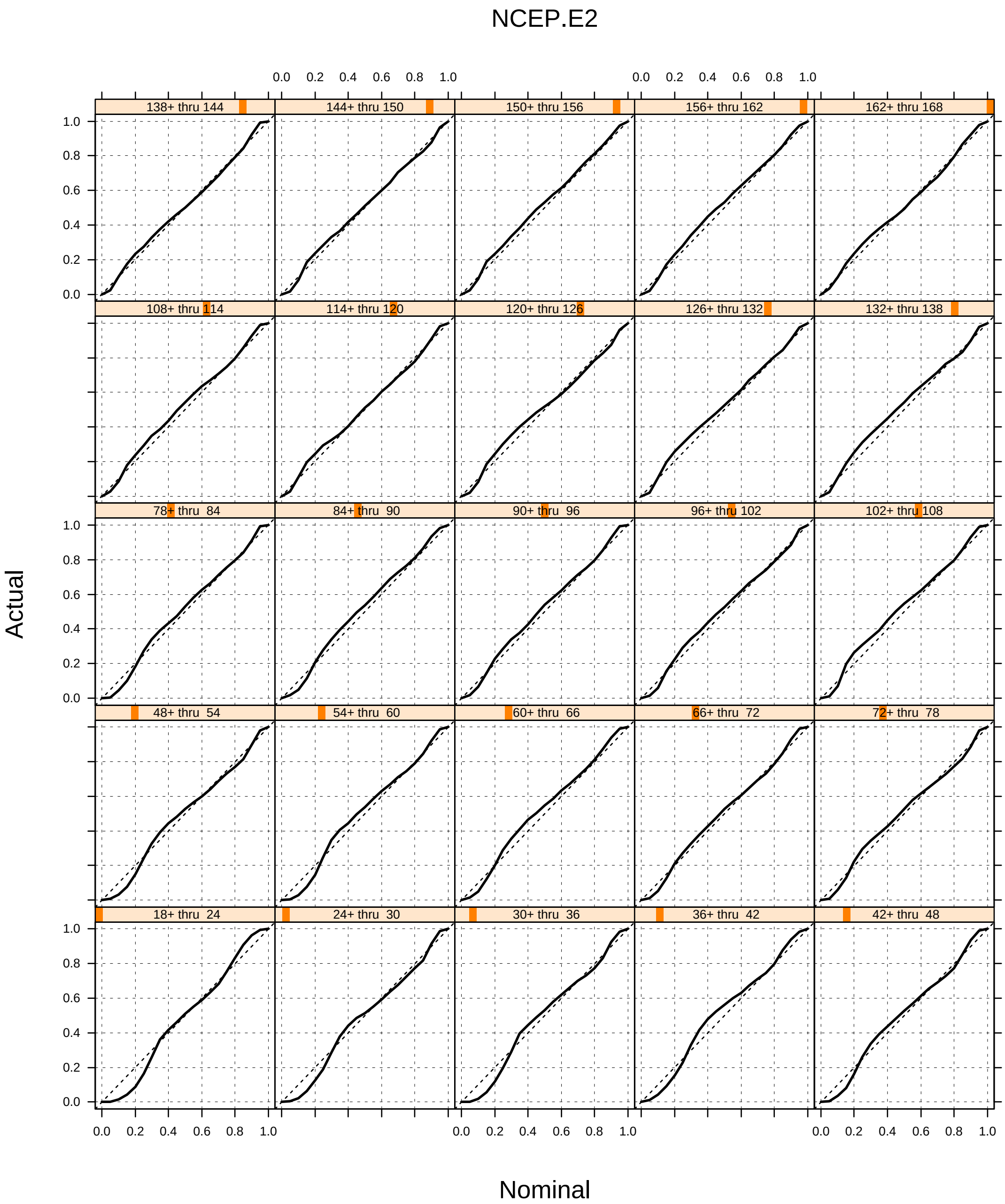

Figure 13: Reliability plots for rerun of E2 using NCEP ensembles.

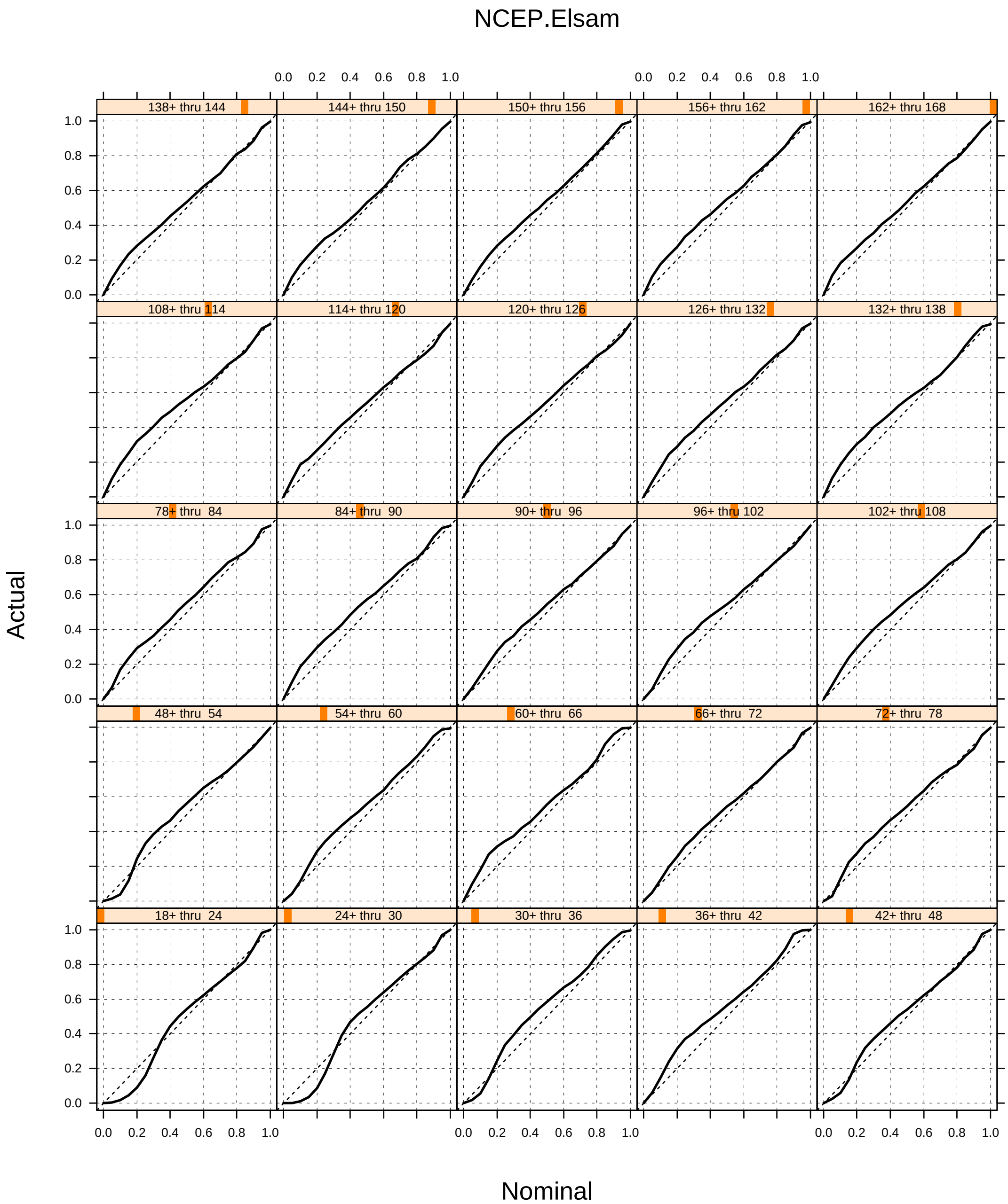

Figure 14: Reliability plots for rerun of Elsam using NCEP ensembles.

B Sharpness and Resulution for a range of quantiles

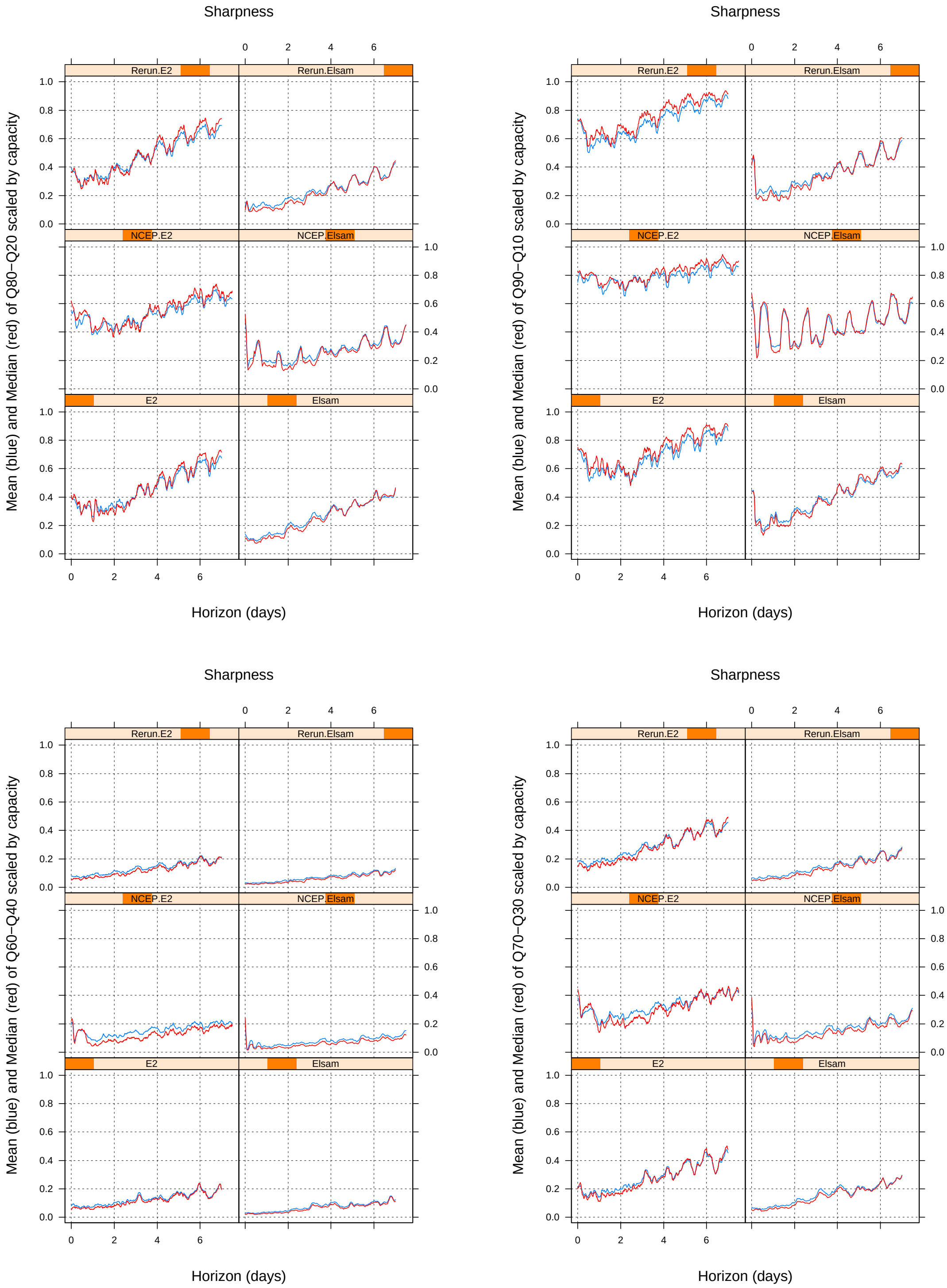

Figure 15: Sharpness evaluated based on differences in quantiles symetric about the median.

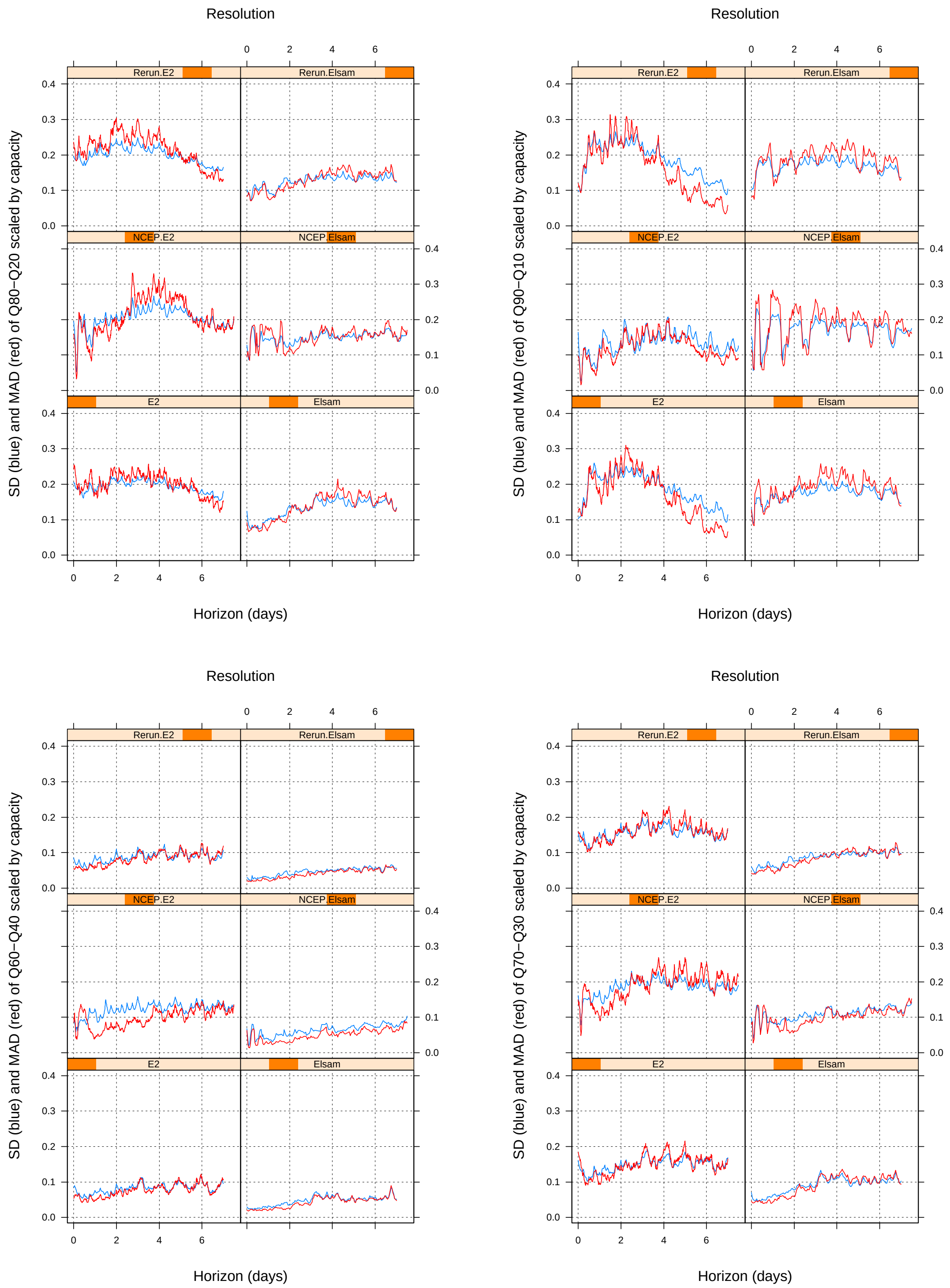

Figure 16: Resolution evaluated based on differences in quantiles symetric about the median.